Archive of SID

مهندسی صنایع و مدیریت شریف (زمستان ۱۳۹۴)
دوره‌ی ۱ - ۳۱، شماره‌ی ۲/۲، ص. ۱۳-۱۸

Original Article

مدل برنامه‌ریزی عدد صحیح برای غربالگری متغیرهای سیستم‌های چندمتغیره با استفاده از سیستم ماهالانوبیس تاگوچی (مطالعه‌ی موردی: خسارت شرکت‌های بیمه)

امیرحسین براهیمی (دانشجوی کارشناسی ارشد)
عبدالله آقایی* (استاد)
دانشکده‌ی مهندسی صنایع، دانشگاه صنعتی خواجه نصیرالدین طوسی

امروزه با رشد سریع تکنولوژی، جمع‌آوری اطلاعات زیاد و مورد نیاز برای تجزیه و تحلیل فرایند در سازمان‌ها ممکن شده است، امّا مهم‌تر از آن امکان به‌کارگیری بهینه و مؤثر این تعداد زیاد داده‌ها در تجارت مدرن است. هدف این مقاله استفاده از سیستم ماهالانوبیس تاگوچی برای حذف اطلاعات اضافی، و نیز تجزیه و تحلیل و تشخیص آنها سیستم‌های چندبعدی است. سیستم ماهالانوبیس تاگوچی از دو بخش کلی تشکیل شده است: ۱. انتخاب متغیرهای مفید؛ ۲. پیش‌بینی و تشخیص. در این سیستم برای غربالگری متغیرها از آرایه‌های متعامد تاگوچی استفاده شده است. در حقیقت نوآوری اصلی این مقاله، استفاده از مفهوم «طبقه‌بندی اشتباه و مدل‌سازی برنامه‌ریزی ریاضی عدد صحیح» به‌جای آرایه‌های متعامد تاگوچی به‌منظور به دست آوردن مجموعه‌یی مفید از متغیرهاست. با این روش دست‌یابی به مجموعه‌یی بهتر در زمانی کوتاه‌تر ممکن می‌شود.

واژگان کلیدی: سیستم‌های چندمتغیره، سیستم ماهالانوبیس تاگوچی، برنامه‌ریزی عدد صحیح.

a.h.barahimi@gmail.com
aaghaie@kntu.ac.ir

۱. مقدمه

سیستم ماهالانوبیس تاگوچی[۱و۲] (MTS)[۱] روشی است برای تجزیه و تحلیل سیستم‌های چندمتغیره که به‌منظور شناسایی و تشخیص الگوی این سیستم‌ها توسط تاگوچی و جوگولوم (۲۰۰۲)[۲] ارائه شد و به دلیل برخورداری از دقت کافی، در صنعت‌های مهم آمریکا و ژاپن و سپس در سرتاسر دنیا به شدت مورد استفاده قرار گرفت. همچنین مقالات بسیاری درخصوص کاربرد این سیستم در سایر بخش‌ها ــ نظیر سیستم بانک‌داری، سیستم دانشگاهی، سیستم بیمه، سیستم تشخیص عیب نرم‌افزارهای رایانه‌یی، سیستم بورس و غیره ــ ارائه شده که خود بیان‌گر اهمیت این موضوع است.

برخی از مسائل مفهومی، عملیاتی و فنی MTS توسط محققین متعددی مورد بحث قرار گرفته است. در مسائل عملیاتی، آنها اعتقاد دارند که استفاده از آزمایشات فاکتوریل کسری و نرخ سیگنال به اغتشاش برای کمینه‌سازی متغیرها (در شرایط استفاده از متغیرهای اصلی) به‌منظور تشخیص مؤثر، بی‌نتیجه است. بررسی آنها اشاره به این دارد که از الگوریتم جست‌وجوی بهتری می‌توان برای تعیین ترکیب بهینه‌ی متغیرها بهره گرفت. روش کاهش ابعاد جدید را می‌توان جایگزین آرایه‌های متعامد و روش طراحی آزمایش برای تعیین ترکیب بهینه‌ی متغیرها ــ که دارای بیشترین نسبت S/N هستند ــ ساخت. جوگولوم و همکارانش اشاره کرده‌اند که روش غربالگری متغیر درصورتی توسعه می‌یابد که مقیاس اندازه‌گیری بهتری نسبت به آرایه‌های متعامد ارائه دهد، و این امر را می‌توان به‌راحتی در تجزیه و تحلیل MTS گنجاند.[۱] محققین تأیید کرده‌اند که الگوریتم‌های بهتری برای کاهش ابعاد می‌توان توسعه داد و OA[۲] ممکن است بهترین استراتژی برای انتخاب زیرمجموعه نباشد.[۳] آنان همچنین با استفاده از برنامه‌ریزی عدد صحیح و حل آن به‌وسیله‌ی الگوریتم PSO [۴] و نیز توسط الگوریتم باینری مورچگان[۵] برای غربالگری متغیرهای مهم روشی ارائه کردند. لذا در این مقاله روشی ساده‌تر و در عین حال دقیق‌تر برای غربالگری متغیرهای مفید ارائه می‌شود. همچنین سعی شده تا با استفاده از روش پیشنهادی انتخاب متغیر، متغیرهای مهم برای سیستم بیمه شناسایی شود. در روش پیشنهادی، از فاصله‌ی ماهالانوبیس (MD)[۳] و ماهیت کاربردی آن به‌منظور تمییز گروه نرمال و غیرنرمال، و نیز برنامه‌ریزی عدد صحیح ریاضی که در این خصوص انجام می‌شود، بدون به‌کارگیری الگوریتم‌های فراابتکاری برای شناسایی متغیرهای مفید استفاده می‌شود.

بخش ۲ این مقاله به توضیح سیستم ماهالانوبیس تاگوچی و ابعاد آن پرداخته و سپس در بخش ۳ اصول مدل‌سازی و فرضیات آن بررسی می‌شود. در بخش ۴ به تشریح چگونگی حل این مدل اختصاص یافته و در بخش ۵ پیاده‌سازی این

* نویسنده مسئول
تاریخ: دریافت ۱۳۹۲/۶/۲۷، اصلاحیه ۱۳۹۳/۳/۲۵، پذیرش ۱۳۹۳/۴/۲۲.

www.SID.ir

Archive of SID

مدل برای داده‌های سیستم بیمه و نیز مقایسه و نقد مطالعات گذشته ذکر شده است. در نهایت بخش ۶ نیز به نتیجه‌گیری اختصاص یافته است.

۲. سیستم ماهالانوبیس تاگوچی[۶-۱۰]

در ابتدا داده‌ها به دو گروه نرمال یا سالم و غیرنرمال یا ناسالم دسته‌بندی می‌شود؛ مثلاً در سیستم تشخیص پزشکی داده‌های نرمال مشخص‌کننده‌ی افراد سالم‌اند، و داده‌های غیرنرمال مشخص‌کننده‌ی افرادی هستند که دارای بیماری مورد نظر هستند. در سیستم بانک‌داری داده‌های نرمال به افرادی اشاره دارند که چک آنها به موقع وصول شده، و داده‌های غیرنرمال افرادی هستند که چک آنها برگشت خورده است. حال با این تقسیم‌بندی می‌توان گفت که سیستم ماهالانوبیس تاگوچی از چهار مرحله برای اجرا تشکیل می‌شود:

الف) ساختار فضای ماهالانوبیس. متغیرهایی که به‌عنوان مشخصه‌های «سالم» تعریف شده‌اند، مشخص می‌شوند. داده‌ها درگروه‌های نرمال یا سالم جمع‌آوری می‌شوند. سپس متغیرهای نرمال استانداردسازی، و برای اجزای نرمال MD محاسبه می‌شود. این مقادیر به‌عنوان «فضای ماهالانوبیس» تعریف، و به‌عنوان چارچوب مرجعی برای مقیاس اندازه‌گیری MTS مورد استفاده قرار می‌گیرد:

$$\mathrm{MD}_j = D^2 = \frac{1}{k} Z_{ij} C^{-1} Z_{ji}^T \qquad (1)$$

n تعداد مشاهدات؛ k تعداد مشخصه‌ها؛ i شماره‌ی مشخصه $(i = 1, 2, ..., k)$؛ j شماره‌ی مشاهده $(j = 1, 2, ..., n)$؛ Z_{ij} مقدار استاندارد شده‌ی مشاهده‌ی X_{ij} (jامین مشاهده از iامین مشخصه‌ی کیفی)؛ C ماتریس همبستگی؛ C^{-1} معکوس ماتریس همبستگی؛ MD_j فاصله‌ی ماهالانوبیس jامین مشاهده.

ب) اعتبارسنجی MS. در این مرحله فاصله‌ی ماهالانوبیس برای داده‌های گروه غیرنرمال محاسبه می‌شود، به این صورت که برای استانداردسازی آنها از میانگین و انحراف معیار داده‌های گروه نرمال استفاده می‌شود و برای محاسبه فاصله‌ی ماهالانوبیس هرکدام از این مشاهدات، ماتریس همبستگی گروه نرمال کاربرد دارد. دلیل این کار مشخص کردن فاصله‌ی هرکدام از مشاهدات از گروه نرمال است. طبق MTS این کار زمانی خوب عمل می‌کند که فاصله‌ی ماهالانوبیس برای گروه غیرنرمال بزرگ‌تر از فاصله‌ی ماهالانوبیس برای گروه نرمال باشد.

ج) انتخاب متغیرهای مفید (بهینه‌سازی). در بسیاری از موارد به دست آوردن اطلاعات بسیار وقت‌گیر و هزینه‌زاست، و نیز تجزیه و تحلیل حجم بسیار وسیعی از اطلاعات نیز بسیار طاقت فرساست؛ لذا با غربالگری داده‌ها می‌توان در وقت، هزینه و... صرفه‌جویی کرد که خود نوعی بهینه‌سازی است. در این مرحله با استفاده از طراحی آزمایش تاگوچی و آرایه‌های متعامد، از بین متغیرهای موجود متغیرهایی را به دست می‌آورند، که این روش چندان قابل اعتماد نیست. به‌همین دلیل تمرکز این مقاله بر این قسمت است تا بتوان با دقت بیشتری این متغیرها را غربالگری کرد.

د) تشخیص و پیش‌بینی. این مرحله به تشخیص آینده و پیش‌بینی با مقیاس MTS براساس متغیرهای مفید اختصاص دارد. بسته به مقدار MD، اصلاح و اقدامات دیگر انجام شده و پیگیری می‌شود. برای پیش‌بینی و تشخیص نرمال یا غیرنرمال بودن مشاهدات آینده، باید از یک حد آستانه استفاده کرد که یک تابع زیان درجه دو برای محاسبه‌ی حدآستانه‌ی مقدار MD مورد استفاده قرار می‌گیرند، به‌طوری که ضرر ناشی از ۲ نوع خطای طبقه‌بندی متعادل گردند.

۱.۲. نرخ سیگنال به اغتشاش (S/N)

یکی از عواملی که بیشتر درمورد مقایسه‌ی ترکیب‌های به دست آمده استفاده می‌شود «نسبت سیگنال به اغتشاش» (S/N) است که نشان‌گر حساسیت مشخصه‌ی کیفی مورد بررسی به عوامل خارجی اثرگذار و غیرقابل کنترل در یک فرایند کنترل شده است. در هر آزمایش، همواره به دنبال بالاترین نسبت S/N در نتایج هستیم که نشان می‌دهد اثر پارامترهای قابل کنترل بیشتر از اثر پارامترهای غیرقابل کنترل است.[۲] محاسبه‌ی S/N مطابق رابطه‌ی ۲ انجام می‌شود:[۲]

$$\eta = -10 log \left[\frac{1}{t} \sum_{j=1}^{t} \frac{1}{\mathrm{MD}_j^2} \right] \qquad (2)$$

که در آن η معرف نسبت سیگنال به اغتشاش، و t تعداد غیرنرمال‌ها تحت شرایط تعیین شده است.

همانطورکه ذکر شد در این بخش اصول کلی سیستم ماهالانوبیس تاگوچی که عمدتاً در تمامی سیستم‌های کارشده تاکنون کاربرد داشته مورد بررسی قرارگرفت. از اینجا دستاوردهای این مطالعه مورد بحث قرار می‌گیرد.

۳. مدل‌سازی برای غربالگری متغیرهای مفید (روش پیشنهادی)

مدل‌سازی پیشرو با استفاده ازوزن‌های داده شده براساس هزینه‌ی طبقه‌بندی اشتباه، و به‌منظور پرداختن به یک مسئله‌ی عملیاتی مربوط به MTS انجام می‌شود. این مقاله برکاهش ابعاد که از شاخصه‌های اصلی یک زیرمجموعه‌ی ساده از مشخصه‌هاست نظارت دارد. مسئله‌ی انتخاب مشخصه، به‌عنوان برنامه‌ریزی عدد صحیح باینری مدل شده و از طریق انتخاب وزن مناسب برای طبقه‌بندی اشتباه در محاسبات، دیدگاه پیشنهادی تصمیم‌گیرنده را برای تعیین زیرمجموعه‌ی مورد نظر از مشخصه‌ها یاری می‌کند. مقدار MD از دو قسمت مختلف برای تشخیص طبقه‌بندی اشتباه استفاده می‌شود. تعریف طبقه‌بندی اشتباه برای پوشش دادن به هر دو طبقه‌بندی اشتباه داده‌های ناسالم به‌جای سالم و بالعکس از طریق یک دیدگاه مجموع وزنی کامل می‌شود.

۱.۳. تعریف معیار غربالگری[۴و۵]

در راستای ارائه‌ی یک مقیاس اندازه‌گیری، داده‌ها به دو گروه «سالم» و «ناسالم» طبقه‌بندی می‌شوند که در این طبقه‌بندی باید کم‌ترین میزان خطا در طبقه‌بندی وجود داشته باشد؛ یعنی مقدار داده‌های سالم که به اشتباه در گروه ناسالم قرار گرفته‌اند و بالعکس باید به حداقل برسد. از این‌رو مجموعه‌ی مفید متغیرها، به مجموعه‌یی اطلاق می‌شود که در ترکیب ساخت آن‌ها از کم‌ترین «طبقه‌بندی اشتباه» استفاده شده است. از این مفهوم نظری به‌عنوان معیاری برای غربالگری متغیرها استفاده می‌شود. در ادامه، چگونگی تشخیص طبقه‌بندی اشتباه با استفاده از مقدار MD، برای مشاهدات «سالم» و «ناسالم» توضیح داده می‌شود.

درکتاب سیستم ماهالانوبیس تاگوچی آمده است که مقدار MD اجزای غیرنرمال از اجزای نرمال بزرگ‌تر است و این جزء اصول اولیه‌ی MTS است. با استفاده از این مفهوم می‌توان حدود بالا و پایینی برای مشاهدات سالم و ناسالم تشکیل داد که مرزی برای تفکیک مشاهدات از یکدیگر است. «طبقه‌بندی اشتباه» هشداری است

www.SID.ir

Archive of SID

ناشی از شناسایی نادرست یک مشاهده در یک رده یا طبقه، زمانی که در واقع به یک گروه متفاوت تعلق دارد. در MTS دو نوع طبقه‌بندی اشتباه وجود دارد: ۱. طبقه‌بندی مشاهدات سالم به‌عنوان ناسالم؛ ۲. طبقه‌بندی مشاهدات ناسالم به‌عنوان سالم. عواقب ناشی از دو نوع مختلف طبقه‌بندی اشتباه به‌شدت به سیستم تحت بررسی بستگی دارد و ممکن است متفاوت باشد. هزینه‌های تحمیلی به سازمان برای دو نوع طبقه‌بندی اشتباه ممکن است یکسان نباشد. تعداد کل طبقه‌بندی اشتباه را می‌توان به‌سادگی با جمع‌کردن دو نوع طبقه‌بندی اشتباه به دست آورد. با این‌حال برای اهداف عملی مقدار «طبقه‌بندی اشتباه» به‌وسیله‌ی جمع وزنی این دو نوع طبقه‌بندی اشتباه محاسبه می‌شود. به این ترتیب یک تابع برای اندازه‌گیری طبقه‌بندی اشتباه کل به‌صورت زیر حاصل می‌شود:

$$\mathrm{TWFM} = f_1(x) = \alpha\frac{Ne_1}{N_1} + \beta\frac{Ne_2}{N_2} \quad (۳)$$

که در آن، Ne_1 تعداد مشاهدات سالم طبقه‌بندی شده در گروه ناسالم؛ Ne_2 تعداد مشاهدات ناسالم طبقه‌بندی شده در گروه سالم؛ N_1 تعداد مشاهدات گروه سالم؛ N_2 تعداد مشاهدات گروه ناسالم؛ α و β به ترتیب زیان‌های ناشی از طبقه‌بندی اشتباه این دو گروه هستند که ممکن است از نوع هزینه باشند؛ یعنی اگر هزینه‌ی ناشی از طبقه‌بندی اشتباه گروه‌های نرمال را با C_1 و هزینه‌ی ناشی از طبقه‌بندی اشتباه گروه‌های غیرنرمال با C_2 نمایش داده شود، α و β چنین محاسبه می‌شود:

$$\alpha = \frac{C_1}{C_1 + C_2} \quad (۴)$$

$$\beta = \frac{C_2}{C_1 + C_2} \quad (۵)$$

برای اندازه‌گیری مقدار Ne_1 و Ne_2 از روش پیشنهادی ارائه شده در مقاله‌ی آویشه و مایتی استفاده نمی‌شود زیرا در این روش، در سیستم‌های مختلف با توجه به ماهیت‌شان ممکن است اختلاف چندانی بین MD در مشاهدات نرمال با غیر نرمال نباشد و نیز ممکن است داده‌های پرت وجود داشته باشد؛ به‌همین دلیل این دیدگاه برای طبقه‌بندی اشتباه به مشکل برمی‌خورد. از این رو در این روش پیشنهادی، یک حد پایین برای مشاهدات غیرنرمال تعریف می‌شود تا هرکدام از مشاهدات غیرنرمال از این حد کم‌تر شدند یک طبقه‌بندی اشتباه محسوب شوند، یعنی این مشاهده در اصل نرمال بوده ولی به اشتباه جزء غیرنرمال‌ها دسته‌بندی شده است. همچنین برای مشاهدات نرمال نیز یک حد بالا در نظر گرفته می‌شود تا هرکدام از این مشاهدات نرمال از این حد بیشتر شدند، جزء طبقه بندی اشتباه محسوب شوند و در اصل این مشاهده غیرنرمال بوده که به اشتباه در مشاهدات نرمال طبقه‌بندی شده است. این حدود بالا و پایین عبارت‌اند از:

$$Lower\ bound = Mmd_N + (ra/r) * Diffmean + Std_N \quad (۶)$$

$$Upper\ bound = Mmd_{AN} - (ra/r) * Diffmean - Std_{AN} \quad (۷)$$

که در آنها، Mmd_N میانگین فاصله‌ی ماهالانوبیس گروه نرمال؛ Mmd_{AN} میانگین فاصله‌ی ماهالانوبیس گروه غیرنرمال؛ Std_N انحراف استاندارد فاصله‌ی ماهالانوبیس گروه نرمال؛ Std_{AN} انحراف استاندارد فاصله‌ی ماهالانوبیس گروه غیرنرمال؛ $Diffmean$ اختلاف میانگین فاصله‌ی ماهالانوبیس گروه نرمال و غیر نرمال؛ r تعداد مشاهدات گروه نرمال؛ ra تعداد مشاهدات گروه غیرنرمال.

حال تعداد طبقه‌بندی‌های اشتباه گروه نرمال را با Ne_1 و تعداد طبقه‌بندی‌های اشتباه گروه غیرنرمال را با Ne_2 نمایش می‌دهیم. با این تفاسیر، باید از تابع TWFM برای کمینه‌سازی خطا بهره گرفت. نکته‌یی که باید در بهینه‌سازی مسئله در نظر گرفت این است که باید تا جایی که به مسئله آسیب نرسد تعداد متغیرهای مفید را کاهش داد و آن را چنین فرموله کرد:

$$f_2(x) = \frac{P_{Select}}{P} \quad (۸)$$

که در آن P تعداد کل متغیرها و P_{Select} تعداد متغیرهای انتخاب شده است. در نتیجه فرمول‌بندی کلی تابع عبارت خواهد بود از:

$$F(x) = W_1 * f_1(x) + W_2 * f_2(x) \quad (۹)$$

W_1 و W_2 وزن‌های این دو تابع هستند و به سیاست تصمیم‌گیرنده بستگی دارد. محدودیت‌های این مسئله نیز عبارت است از:

$$\sum_{i=1}^{p} x_i \le p \quad (۱۰)$$

$$\sum_{i=1}^{p} x_i = p_{selected} \quad (۱۱)$$

$$f_1(X) \le {f_1}^{max} \quad (۱۲)$$

فرض کنید $X = (x_1, x_2, ..., x_p)^t$ یک بردار P بعدی است که در آن:

$$x_i = \begin{cases} 0 & \text{اگر متغیر } i \text{ انتخاب شود} \\ 1 & \text{اگر متغیر } j \text{ انتخاب شود} \end{cases}$$

بدیهی است که هر دو تابع $f_1(x)$ و $f_2(x)$ اعداد مثبت بین ۰ و ۱ هستند. این تابع هدف با توجه به محدودیت‌های ۱۰ الی ۱۲ بهینه‌سازی می‌شود. فرمول‌بندی به دست آمده یک برنامه‌ریزی باینری عدد صحیح است که در آن محدودیت اول تصریح می‌کند که مجموع متغیرهای تصمیم باید کوچک‌تر یا مساوی تعداد متغیرها در کل مجموعه داده باشد. (هنگامی که زیرمجموعه‌یی از متغیرها برای آماده‌سازی مقیاس اندازه‌گیری MTS با استفاده از این روش انتخاب شد، مجموع متغیرهای تصمیم به‌شدت کم‌تر از تعداد اولیه‌ی متغیرها می‌شود.) معادله‌ی دوم اندازه‌ی یک زیرمجموعه از متغیرها را بیان می‌کند. محدودیت آخر تمایل تصمیم‌گیرنده را برای قبول صرفاً زیرمجموعه‌ی مشخصه‌هایی جلب می‌کند که وزن کل ایجادشده‌ی طبقه‌بندی اشتباه کوچک‌تر یا مساوی تمام متغیرهای تولید شده است. در موارد عملی، تصمیم‌گیران از ارائه‌ی یک مقیاس اندازه‌گیری خودداری می‌کنند، که این نسبت به حالتی که با تمام متغیرهای اصلی ایجاد شده است، کم‌تر تبعیض‌آمیز است و باعث می‌شود تصمیم‌گیرنده ${f_1}^{max}$ را اتخاذ می‌کند و با متغیرهای اولیه‌ی ارائه شده توسط وی، محاسبات انجام می‌شود.

۴. روش حل این مدل

این مدل را به روش‌های متفاوتی می‌توان حل کرد اما دقیق‌ترین روش بررسی تمامی حالات ممکن است. بنابراین مسئله با استفاده از روش بررسی تمام نقاط ممکن با نرم‌افزار متلب R۲۰۱۱a و CPU Core(TM) i۵-۲۴۱۰M ۲٫۳۰GHz برای گروه‌هایی با متغیرهای متفاوت حل شد که نتایج زمانی جدول ۱ را در برداشته است.

حال اگر برآوردی از مشاهدات آتی لازم باشد، می‌توان با استفاده از نرم‌افزار MiniTab معادله‌یی براساس تعداد متغیرها به دست آورد:

$$Y = ۰٫۷۸۳۲ - ۰٫۴۷۲ * X + ۰٫۰۸۲۳۵ * X^2 \quad (۱۳)$$

www.SID.ir

Archive of SID

جدول ۱. نتایج زمانی حاصل از حل مدل به‌صورت کامل.

تعداد متغیرها	تعداد مشاهدات	زمان حل
۳	۲۳۰۰ مشاهده نرمال و ۹۷۰ مشاهده غیرنرمال	۰٫۱۱
۵	۲۳۰۰ مشاهده نرمال و ۹۷۰ مشاهده غیرنرمال	۰٫۴۸
۷	۲۳۰۰ مشاهده نرمال و ۹۷۰ مشاهده غیرنرمال	۱٫۵۱
۹	۲۳۰۰ مشاهده نرمال و ۹۷۰ مشاهده غیرنرمال	۳٫۲۱
۱۱	۲۳۰۰ مشاهده نرمال و ۹۷۰ مشاهده غیرنرمال	۵٫۵۵

که در آن X تعداد مشخصه‌های موجود در مدل و Y زمان صرف شده برای حل مدل است. از آنجا که زمان حل نسبت به تعداد متغیرها رابطه‌ی خطی ندارد و به‌صورت توابع نمایی یا چندمرتبه‌یی است، بهترین منحنی برازش شده که برای رابطه‌ی ۱۳ دارای خطای کم‌تری است با تشخیص نرم‌افزار MiniTab به‌صورت فوق است.

با استفاده از معادله‌ی ارائه شده، برای داده‌های عملی که در تمام مقاله‌ها مورد بررسی قرارگرفته و همچنین درکتاب ماهالانوبیس تاگوچی، بیشترین تعداد متغیرهای موجود ۴۰ و بیشترین تعداد مشاهدات ۴۰۰ است؛ با این وجود حداکثر زمان حل آن ۱۱۴ ثانیه برآورد می‌شود و درنتیجه به نظر می‌رسد که نیازی به حل از طریق روش‌های فراابتکاری نباشد.

۵. پیاده‌سازی روش پیشنهادی روی سیستم بیمه بدنه‌ی خودرو[۱۱]

شرکت‌های بیمه یکی از مهم‌ترین ارگان‌های خدماتی هستند که امروزه جایگاه مهمی در این بین دارند. بیمه با ایجاد امنیت خاطر و آرامش روحی، نگرانی ناشی از حوادث ناگهانی و پیش‌بینی نشده را در اذهان کمرنگ کرده و از این طریق به روان شدن چرخه‌ی کار و زندگی کمک می‌کند.

هنگامی که در سیستمی عوامل مؤثر زیادی وجود داشته باشد و این عوامل مستقل از هم نباشند و بر هم اثر بگذارند، تجزیه و تحلیل سیستم پیچیده می‌شود. در چنین شرایطی با استفاده از تحلیل‌های چندمتغیره می‌توان با شناسایی عوامل مهم‌تر و در نظرگرفتن اثر تمام عوامل بر هم، سیستم را ساده‌تر و قابل فهم‌تر کرد. روش MTS یک روش تحلیل چندمتغیره‌ی مناسب برای کاهش عوامل تأثیرگذار و شناسایی عوامل اصلی و مهم‌تر است. این قسمت با استفاده از توانایی روش پیشنهادی، عوامل اثرگذار در تصادفات را شناسایی می‌کند.

۱.۵. گام‌های اساسی پیاده سازی روش پیشنهادی

گام ۱. تعریف متغیرها: براساس نظر کارشناسان بیمه‌ی بدنه‌ی خودرو هفت عامل اصلی به‌عنوان عوامل تأثیرگذار در خسارت وارده به یک خودرو شناسایی شده‌اند. این عوامل به همراه دامنه‌ی عوامل در جدول ۲ معرفی شده است. (داده‌ها مربوط به سال ۱۳۹۰ است.)

گام ۲. تعریف گروه نرمال: برای این منظور از داده‌های واقعی مشتریانی که در سال ۱۳۹۰ خودروی خود را در یکی از شرکت‌ها بیمه کرده‌اند، استفاده شده است. از آنجا که کارشناسان عامل سابقه‌ی بیمه را به‌عنوان یک متغیر مهم در این زمینه معرفی کرده‌اند، مشتریانی در نظرگرفته شده‌اند که خودروی آنها سواری و با استفاده‌ی شخصی است. گروه نرمال مشتریانی هستند که در این سال به شرکت سود رسانده‌اند و گروه غیرنرمال مشتریانی هستند که طی سال ۱۳۹۰ به شرکت ضرر رسانده‌اند. فرض بر این است که افرادی که خودرو را بیمه کرده‌اند از خودرو استفاده کرده‌اند.

گام ۳. تولید یک مقیاس داده ازگروه نرمال برای متغیرهای انتخاب شده: از بین مشتری‌های موجود در شرکت بیمه با اطلاعات مورد نظر، ۷۵۰ مشاهده‌ی نرمال برای تمام متغیرهای تعریف شده به‌صورت تصادفی نمونه‌گیری می‌شوند.

گام ۴. جمع‌آوری مشاهدات خارج ازگروه نرمال: دراینجا ۲۵۰ مشاهده‌ی غیرنرمال به‌طور تصادفی نمونه‌گیری می‌شود.

گام ۵. کدکردن مسئله با استفاده از مدل‌سازی انجام شده در بخش قبل: هزینه‌ی طبقه‌بندی اشتباه مشاهدات نرمال به‌جای غیرنرمال با نظر تصمیم‌گیرنده ۹۰ $= C_1$ و هزینه‌ی طبقه بندی‌اشتباه مشاهدات غیرنرمال به‌جای نرمال ۳۰ $= C_2$ است. درنتیجه α و β به ترتیب ۰٫۷۵ و ۰٫۲۵ است. نتایج این مدل‌سازی در جدول ۳ با وزن‌های مختلف ارئه شده است.

در جدول ۳ برای مقایسه‌ی کارآیی مدل با مدل‌سازی مایتی و همکارانش، و همچنین روش طراحی آزمایش، از نرخ S/N استفاده می‌شود. چنان‌که ملاحظه می‌کنید با توجه به نظر تاگوچی درکتاب سیستم ماهالانوبیس تاگوچی ــ مبنی بر این که هرچه نرخ S/N بیشتر باشد، آن ترکیب کارآیی بهتری دارد ــ نرخ S/N برای روش پیشنهادی با وزن‌های داده شده نسبت به روش طراحی آزمایش تاگوچی (که از جدول آرایه‌های متعامد برای ۷ مشخصه در دو سطح است) و مدل مورد استفاده‌ی مایتی و همکارانش که در آن وزن $W = (۰٫۹, ۰٫۱)$ در نظرگرفته شده، بهتر است.

۲.۵. تحلیل حساسیت نسبت به پارامترهای α و β

با توجه به جدول ۴، برای وزن‌های $W = (۰٫۸, ۰٫۲)$ تغییرات α و β در نتایج این مسئله بی‌تأثیر است و برای وزن‌های $W = (۰٫۹, ۰٫۱)$ نتایج حاکی از این است که در صورت برابری α و β از تعداد متغیرهای مفید به دست آمده، کم می‌شود و برای وزن‌های $W = (۰٫۹, ۰٫۱)$ تغییرات α و β در نتایج تأثیر ندارد و در حالت کلی می‌توان گفت در این بررسی موردی وزن‌های α و β تأثیری ندارند، البته نکته‌یی که در اینجا مطرح است صدق کردن این نتایج فقط درمورد این مسئله است و ممکن است در مسائل دیگر این پارامترها مؤثر باشند؛ به همین دلیل در مدل ارائه شده α و β گنجانیده شده است.

۶. نتیجه‌گیری

چنان‌که گفته شد، غربالگری متغیرها در دنیای واقعی باعث صرفه‌جویی در هزینه و زمان می‌شود و افزون بر این، در سیستم‌های پیش‌بینی چندمتغیره که یکی از کاراترین آنها سیستم ماهالانوبیس تاگوچی است امری بسیار ضروری است. با توجه به مقایسه‌های انجام شده در قسمت قبل، این نکته مشاهده شد که با اجرای این مدل برای سیستم بیمه، مشخصه‌ی اول (سن) با تغییر وزن‌های مدل گاهاً حذف و اضافه می‌شود. یعنی این متغیر در این مسئله خیلی نسبت به پارامترهای مدل حساسیت نشان می‌دهد ولی دو متغیر دیگر یعنی جنسیت و عمر ماشین، در تمامی ترکیبات

www.SID.ir

Archive of SID

جدول ۲. تعریف متغیرها و دامنه‌ی آنها.

متغیر	اندیس	دامنه
سن	x_1	۷۷-۱۸
جنسیت	x_2	۱-۰ (مرد و زن)
شهر	x_3	A، B، C، D و E (پنج شهر پر جمعیت که نام آنها نزد نویسندگان محفوظ است)
نوع خودرو	x_4	۱۰ ماشین پر طرفدار $m_1, m_2, m_3, m_4, m_5, m_6, m_7, m_8, m_9, m_{10}$
عمر ماشین	x_5	۲۰-۱
سابقه تصادف	x_6	۱-۰
ارزش خودرو	x_7	۳۵۰۰۰۰۰۰۰ ریال - ۳۲۰۰۰۰۰۰ ریال

جدول ۳. نتایج حل مسئله و مقایسات.

مقدار تابع هدف	S/N	$X = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7\}$	وزن‌ها $W = \{w_1, w_2\}$
۰٫۱۲۱۱	۱۱٫۳۲۶۰	$X = \{0, 1, 0, 0, 1, 0, 0\}$	(۰٫۸, ۰٫۲)
۰٫۱۰۰۶	۱۰٫۸۷۸۸	$X = \{1, 1, 0, 0, 1, 0, 0\}$	(۰٫۹, ۰٫۱)
۰٫۰۹۰۳	۱۰٫۸۷۸۸	$X = \{1, 1, 0, 0, 1, 0, 0\}$	(۰٫۹۵, ۰٫۰۵)
۰٫۰۸۶۴	۸٫۷۸۴۶	$X = \{1, 1, 0, 0, 1, 1, 0\}$	(۰٫۹۷۵, ۰٫۰۲۵)
—	۹٫۷۸۱۸	$X = \{0, 1, 0, 0, 1, 0, 1\}$	با مدل مایتی
—	۹٫۷۸۵۰	$X = \{0, 1, 0, 1, 1, 0, 0\}$	با استفاده از طراحی آزمایش‌ها

جدول ۴. نتایج تحلیل حساسیت نسبت به α, β.

مقدار تابع هدف	$X = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7\}$	α, β	$W = \{w_1, w_2\}$
۰٫۲۲۹۹	$X = \{0, 1, 0, 0, 1, 0, 0\}$	(۰٫۲۵, ۰٫۷۵)	
۰٫۱۷۵۵	$X = \{0, 1, 0, 0, 1, 0, 0\}$	(۰٫۵, ۰٫۵)	(۰٫۸, ۰٫۲)
۰٫۱۲۱۱	$X = \{0, 1, 0, 0, 1, 0, 0\}$	(۰٫۷۵, ۰٫۲۵)	
۰٫۲۲۱۱	$X = \{1, 1, 0, 0, 1, 0, 0\}$	(۰٫۲۵, ۰٫۷۵)	
۰٫۱۶۱۸	$X = \{0, 1, 0, 0, 1, 0, 0\}$	(۰٫۵, ۰٫۵)	(۰٫۹, ۰٫۱)
۰٫۱۰۰۶	$X = \{1, 1, 0, 0, 1, 0, 0\}$	(۰٫۷۵, ۰٫۲۵)	
۰٫۲۰۹۵	$X = \{1, 1, 0, 0, 1, 0, 0\}$	(۰٫۲۵, ۰٫۷۵)	
۰٫۱۵۰۶	$X = \{1, 1, 0, 0, 1, 0, 0\}$	(۰٫۵, ۰٫۵)	(۰٫۹۵, ۰٫۰۵)
۰٫۰۹۰۳	$X = \{1, 1, 0, 0, 1, 0, 0\}$	(۰٫۷۵, ۰٫۲۵)	

وجود دارند. به‌همین دلیل عمر ماشین، جنسیت و سن متغیرهای مفیدی هستند که ارائه می‌شود.

با استفاده از مدل پیشنهادی تعداد متغیرها به کم‌ترین مقدار ممکن (۳ از ۷) و همچنین نرخ S/N نیز در بین آنها بیشترین مقدار ممکن را دارد. این امر نمایانگر کارایی این مدل دست کم درمورد سیستم بیمه است. البته روش دیگری نیز برای مقایسه‌ی کارایی مدل وجود دارد و آن هم استفاده از این ترکیب برای پیش‌بینی و تشخیص سیستم است که این مقایسه به حوزه‌ی تحقیقات آتی منتقل می‌شود.

به‌طورکلی انتخاب زیرمجموعه‌ی بهینه از متغیرها علاوه بر سیستم ماهالانوبیس تاگوچی، در زمینه‌های دیگر ازجمله الگوریتم شبکه عصبی، رگرسیون و غیره کاربرد دارد که برای این منظور از غربالگری برای بهینه‌سازی و افزایش سرعت اجراها استفاده می‌شود.

پیاده‌سازی سیستم ماهالانوبیس تاگوچی برای سیستم بیمه باعث می‌شود تا شرکت‌های بیمه مشتریانی راکه خسارت بالایی به شرکت وارد می‌کنند شناسایی کنند و با پیش‌بینی تقاضاهای این شرکت که از نوع خسارت است تمهیدات و سیاست‌های مناسب را برای این موضوع اعمال کند.

www.SID.ir

Archive of SID

پانوشت‌ها

1. Mahalanobis taguchy system
2. Orthogonal Array
3. Mahalanobis Distance

منابع (References)

1. Jugulum, R., Taguchi, G., Taguchi, S. and Wilkins, J.O. "Discussion of a review and analysis of Mahalanobis-Taguchi system", *Technometrics,* **45**(1), pp. 16-21 (2003).
2. Taguchi, G., Jugulum, R., *The Mahalanobis-Taguchi Strategy*, John Wiley & Sons, New York (2002).
3. Variyath, A.M. and Abraham, B. "Discussion", *Technometricse*, **45**(1), pp. 22-24 (2003).
4. Pal, A. and Maiti, J. "Development of a hybrid methodology for dimensionality reduction in Mahalanobis–Taguchi system using Mahalanobis distance and binary particle swarm optimization", *Expert Systems with Applications,* **37**(2), pp. 1286-1293 (2010).
5. Reséndiz, E., Moncayo-Martínez, L.A. "Binary Ant colony optimization applied to variable screening in the Mahalanobis–Taguchi System", Expert Systems with Applications, pp. 634-637 (2012).
6. Taguchi, G. and Jugulum, R., *The Mahalanobis-Taguchi System*, Edition 1 , R.R.Donnelley & Sons, New York (2001).
7. Woodall, W.H., Koudelik, R., Tsui, K.L., Kim, S.B., Stoumbos, Z.G. and Carvounis, C.P. "A review and analysis of the mahalanobis—Taguchi System", *Technometrics*, **45**(1), pp. 1-15 (2003).
8. Mahalakshmi, P. and Ganesan, K. "Mahalanobis Taguchi system based criteria selection for shrimp aquaculture development", *Computers and Electronics*, **65**(2), pp. 192-197 (2009).
9. Mohan, D., Saygin, C. and Sarangapani, J. "Real-time detection of grip length deviation during pull-type fastening: A Mahalanobis–Taguchi system (MTS)-based approach", *The International Journal of Advanced Manufacturing Technology*, **39**(9-10), pp. 995-1008 (2008).
10. Liparas, D., Angelis, L. and Feldt, R. "Applying the Mahalanobis-Taguchi strategy for software defect diagnosis", *Automated Software Engineering*, **19**(2), pp. 141-165 (2012).
11. Abbasi, S.E., Aaghaie, A. and Fazlali, M. "Applying Mahalanobis –Tagouchi system in detection of high risk customers: A case-based study in an insurance company", *Journal of Industrial Engineering,* pp. 1-12 (2011).

www.SID.ir